DOCUMENT RESUME

ED 109 250 TM 004 714

AUTHOR Sewell, Alan F.; And Others
TITLE Controlled Multivariate Evaluation of Open Education: Application of a Critical Model.
PUB DATE [Apr 75]
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975); For related documents, see ED 074 144, 090 299 and 093 930

EDRS PRICE MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS Academic Achievement; *Comparative Analysis; *Junior High Schools; Junior High School Students; *Open Education; Program Development; *Program Evaluation; Social Adjustment; Statistical Analysis; Student Characteristics; Student Evaluation; Test Results; *Traditional Schools

ABSTRACT

This paper continues previous reports of a controlled multivariate evaluation of a junior high school open-education program. A new method of estimating program objectives and implementation is presented, together with the nature and degree of obtained student outcomes. Open-program students were found to approve more highly of their learning environment and to enjoy higher self-concepts than the traditional program control students, and at no loss in academic achievement. Studies of several student "types" showed the open program to be superior for underachievers, introverts, and extraverts. The criteria for transition of the program from "experimental" to "adopted" are discussed. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED109250

TM 004 714

- SUMMARY -

Controlled Multivariate Evaluation of Open Education: Application of a Critical Model

Alan F. Sewell, Allan W. Dornseif, and Susan Gross Sturm
Matteson, Illinois School District 162

This report presents and discusses results of second-year evaluation of an open education program conducted at the junior high school level. The report references and continues previously published reports of this project by the same authors (ERIC Documents 074 144, 090 299, 093 930).

The fundamental evaluation design was that of a 2 X 2 X 2 MANOVA, employing as independent variables: Type of Program (open vs. traditional), Grade Level (7th vs. 8th), and Sex (male vs. female). Major dependent variables were scores on the Stanford Achievement Tests (4 scales), Learning Environment Inventory (14 scales), and Piers-Harris Children's Self Concept Scale.

Two types of analyses were conducted: absolute and relative. Absolute analyses simply compared student outcomes as a function of program. The relative analyses compared student outcomes both between programs and within each program as a function of teachers' ratings of the students on four rating dimensions (attitude, knowledge, skills, sociability). Using these ratings, students within each program were classified as "successful" or "unsuccessful" in an attempt to define program differences and factors contributing to "success" within each program.

Data showed the two groups (open and traditional) not to differ significantly in academic achievement. In terms of the Learning Environment Inventory scales, the open group scored higher on Cliqueness, Disorganization, and Democratic, while the traditional group scored higher on Formality, Speed, Favoritism, and Apathy. The two groups did not differ on the Self Concept scale, but the group-by-grade interaction was significant, with the seventh grade open program students achieving the highest mean score.

Analysis of teachers' ratings showed sex + grade to be significantly related to ratings in the open group, such that eighth grade girls were disproportionately "successful" and eighth grade boys were disproprotionately "unsuccessful." In the traditional program, grade level was not a significant contributor to the "success" rating. In both programs, "successful" students significantly exceeded the academic achievements of "unsuccessful" students.

In the open program, "successful" students characterized their program (per LEI scales) as less marked by speed, favoritism, and satisfaction; in the traditional program, "successful" students characterized their program as marked by greater cohesiveness, more enjoyable physical environment, greater goal direction, greater apathy, and greater disorganization; "unsuccessful" students noted more friction, more democracy, and more satisfaction. In both groups, the more "successful" students had higher self-concept scores.

It was concluded that the open program has established a distinctive and characteristic identity, which promotes non-academic objectives without sacrificing academic achievements. Methodological and technical problems of such evaluations are discussed in some detail.

This report will soon be published as an ERIC document.

Controlled Multivariate Evaluation of Open Education: Application of a Critical Model

Alan F. Sewell, Allan W. Dornseif, and Susan Gross Sturm

Matteson, Illinois School District 162

Among the plethora of proposed educational innovations and alternatives of recent years, the concept of "open education" has remained remarkably viable. A steady stream of open education implementations continues throughout the United States and in a number of other lands as well. As noted in a previous paper (Sewell & Dornseif, 1974), while the general theory of open education suggests a movement toward certain educational objectives, in practice it has been most distinctive as a movement away from traditional educational methods. But because implementations of open education have typically been introduced within existing educational systems -- at least in the United States -- a substantial variety of compromises have been effected between the concepts and methods of "pure" open education and the concepts and methods of traditional education.

Only quite recently has a literature of open education evaluation begun to develop; some of this literature has been previously reviewed (Sewell & Dornseif, 1974). Such reviews fail to indicate a clear and consistent pattern of results; inadequacy of the evaluation design is a common flaw, but in addition the question of degree of openness incorporated into any given implementation is typically left unanswered.

In the absence of any significant standardization among the many open education implementations, each must be individually evaluated. While one purpose of the present paper is to report the results of an intensive -- and extensive, progressive and continuous -- evaluation of one specific program, a more

fundamental purpose is to delineate the characteristics and application of a critical evaluation model.

Planning of the educational and evaluation programs began nearly a year prior to the inauguration of this open education program. (This program was soon characterized by the acronym "OSCAR," both in honor of a previous principal and to represent "Open Space for the Conceptualization of Attitudes and Responsibilities"; regardless of its merits, the acronym has persisted, and the open education program will be so identified herein.) Certain fundamental questions have structured and guided the evaluation design and program, and these continue to be pursued: (1) To what extent are the objectives of traditional education met within the open education program? (2) What different or supplementary objectives are met by the open program? (3) Are there types of students for whom either the traditional or open program is more suitable?

It should be noted that these questions focus specifically upon student outcomes, and they clearly imply direct comparisons between outcomes provided within the open education program and those provided by a traditional education program. Further, because multiple objectives are to be considered, the use of multiple measures is implied. The considerations led to the early adoption of a multivariate analysis of variance (MANOVA) evaluation model involving both an experimental (open education) group and a control (traditional education) group.

Specific characteristics of the open classroom, the educational program conducted therein, and student selection procedures have been previously described in detail (Sewell & Dornseif, 1973; Sewell & Dornseif, 1974; Dornseif et al., 1974) and will only be summarized here. The OSCAR (open education) program is conducted in a single, undivided classroom which houses a total of 140 students in a space somewhat larger than that of four traditional classrooms. The students include equal numbers of seventh-graders and eighth-graders, boys and girls -- 35 students in each such subgroup; these students have been randomly

selected from the District's six sending (primary) schools, in order to represent the District's heterogeneity; a control group of similar students, similarly chosen, pursues the traditional, departmentalized program of the junior high school.

During the program's second year (1973-74), the principle of random assignment of students was maintained, but both the open and control groups were further subdivided into a number of subgroups on the basis of characteristics of particular interest; hence both seventh grade groups included students characterized by their sixth grade teachers as "extraverts," "introverts," "academic underachievers," or "likely to do well in an open classroom" (for the open program) or "likely to do well in a traditional program" (for the control group); another subgroup was simply randomly assigned to each program, as in the preceding year. (It should be noted however that each such subgroup will include a very small number of students.)

Students in the control group follow the regular instructional program of the junior high school, moving from classroom to classroom through the day, instructed by teachers neither required nor specifically encouraged to coordinate their instructional programs, and only in a few instances pursuing schedules identical to those of other control group members. OSCAR students, on the other hand, spend the majority of their days in the same classroom, with the same classmates, and instructed by four teachers (Language Arts, Mathematics, Social Science, and Science) and two teacher-aides. Team teaching and close coordination of instructional programs are emphasized; small and flexible instructional groupings and individualization are encouraged.

During the program's first year (1972-73), it should be noted, OSCAR's eighth-graders had previously experienced a year in the junior high school's traditional program, while OSCAR's seventh-graders had entered the open program directly from a non-departmentalized primary school. The following year (1973-

74), of course, eighth-graders had been the first year's seventh-graders and were the first group to have experienced only the open program of the junior high school; these students, then, are of particular interest in evaluation of the open program. While attrition losses were replaced in the OSCAR classroom, they were not included in the evaluation program; hence the data reported herein are derived from eighth-graders who have experienced two full years of the open program and from seventh-graders who have experienced one full year of the open program.

SUMMARY OF FIRST-YEAR RESULTS

Evaluations conducted during the program's first year (1972-73) were largely exploratory in nature, intended primarily to identify rather general areas of evaluation in which specific outcome differences might be sought. Hence during and at the end of that year a large number of different instruments were administered; certain of these were retained for use in the second year's evaluation effort, but others were discontinued. Year-end analyses were conducted in accordance with the project's basic MANOVA design.

First-year results have previously been published in detail (Sewell & Dornseif, 1974); these results will merely be summarized here in order to provide perspective for the second-year findings.

A major concern of the evaluation project was, of course, whether the OSCAR program would adversely affect academic achievement, which, presumably, is the fundamental objective of traditional education. Analyses of Stanford Achievement scores attained by the OSCAR students and the control students satisfactorily dispelled this concern. The OSCAR group significantly out-performed the control group in Social Studies and Science; otherwise the two groups did not differ. In Spelling and in Language, seventh grade OSCAR students surpassed the achievements of other subgroups. Eighth grade control students, however, excelled

in Arithmetic Computation and Arithmetic Concepts. Other multivariate Fs were significant as expected: eighth-graders generally outperformed seventh-graders, and females outperformed males in those variables demonstrating significant differences.

The Bell Adjustment Inventory (slightly amended to allow for the students' age level) showed significant differences in two subscales: Submissiveness-Self-Assertion and Masculinity-Femininity, such that the control students could be described as more self-assertive and more masculine. Seventh-graders of both groups were found to be more self-assertive than eighth-graders. Females of both groups were found to be higher in emotionality, while males were found (fortunately!) to be higher in masculinity.

Of the 18 scales of the California Psychological Inventory, only one was found to differentiate the two groups: Sense of Well-Being; the control group mean was higher than the OSCAR mean.

The (Bell) School Inventory, an attitude-toward-school instrument, did not differentiate the two groups.

The Piers-Harris Children's Self Concept Scale yielded data showing higher self-concepts on the part of the control students.

Rotter's Locus of Control measure did not differentiate the two groups.

"Success" Analyses

On the assumption that traditional and open education programs, resting upon different theoretical foundations, would have different objectives (to some degree at least), an attempt was made to evaluate each program in terms of its objectives. Unfortunately, repeated attempts to secure behavioral objectives from instructional personnel were unsuccessful; and evaluation efforts were redirected toward differentiating between the objectives of both programs. A line of reasoning was adopted which held that a teacher's evaluation of a

student's success will be a function of the teacher's objectives and of the student's attainment of those objectives. Hence study of the teacher's more highly evaluated students, in comparison to those receiving lower evaluations, should reveal characteristics which contribute to the teacher's objectives.

Each of the four OSCAR teachers and four teachers of the control students were asked to rate their respective students in each of four rating dimensions (each of which was behaviorally defined): attitude, knowledge, skills, and sociability. Each OSCAR teacher's ratings were subsequently converted into an individual z-distribution; each student's four ratings provided by that teacher were then z-converted; the four teachers' z-converted ratings were then summed, and in this manner each student was assigned a mean z-rating based upon 16 individual ratings. The same procedure was followed with respect to control students' ratings by their teachers.

Students achieving mean z-ratings above the median for their group were (somewhat arbitrarily) classed as "successful" students, while those below the median were considered "unsuccessful." Multivariate analyses then compared "successful" to "unsuccessful" students in each group and -- more importantly -- attempted to detect significantly different variables as a way of defining "success" in the two groups.

Quite a number of variables were found to differentiate "successful" from "unsuccessful" students in both groups, and, in general, there were few differences in the patterns for both groups: "successful" OSCAR students were found to have a higher sense of communality and a more internalized locus of control than "unsuccessful" students in that group, while neither of these distinctions held for the control group. "Successful" OSCAR students were found to differ from their counterparts in the control group only in the OSCAR students' lower self-concepts and lower sense of well-being. "Unsuccessful" OSCAR students differed from their control counterparts only in their lower

sense of community. In any event, "successful" students in both groups were most clearly differentiated from their "unsuccessful" classmates on the bases of academic achievement variables.

These patterns rather strongly suggested that objectives of the OSCAR program differed little, if any, from those of the traditional program. This conclusion was supported by an additional finding: that intercorrelations of the OSCAR teachers' ratings of their students were not particularly impressive, averaging approximately .50. A team-teaching approach would seem to imply more highly intercorrelated ratings, and a similar implication would seem to be warranted in the case of a highly individualized program.

Further analyses of the teachers' use of student ratings revealed the existence of certain apparently biasing factors, such that within the OSCAR group a "successful" student was most likely to be a seventh grade girl; neither grade level nor sex was found to be significantly predictive of "success" in the control group.

Conclusions and Subsequent Developments

These findings led to an inevitable conclusion that in practice the OSCAR program had not differentiated itself from traditional education: the teachers' emphases continued to rest almost totally upon academic achievement. Problems of classroom management may well have contributed to the emphases, since teachers' comments frequently concerned discipline, order, and noise; such difficulties are also suggested by the ratings biases in favor of those who would, presumably, be the most cooperative and orderly students.

The final report of this first-year evaluation noted (Sewell & Dornseif, 1974, p. 23-24): "It is to be expected that as their experience accumulates the OSCAR teachers will develop coping strategies which will enable them to focus more effectively upon their students' non-cognitive development; that

is, "success" will be redefined to expand concepts of responsibility and sociability and to de-emphasize concepts of order and discipline. The first year of this program has clearly demonstrated that academic standards can be maintained. With the teachers' increasing confidence and competence, the potential contributions of open education to the student's personal and social development can more meaningfully be assessed."

The results, conclusions, and implications of this evaluation were subsequently presented to and discussed with the OSCAR teachers; they were encouraged to develop appropriate objectives and strategies. One of the four OSCAR teachers elected not to continue in the program and was replaced by another volunteer.

SECOND-YEAR EVALUATION

Students who had been seventh-graders during the first year of this study were, of course, eighth-graders during the second year (1973-74); and by the conclusion of that year, these students had experienced two full years in their respective programs (OSCAR or traditional). Neither group, it should be remembered, had had any experience with the other group's program. The OSCAR eighth-graders, hence, were free from the possible contamination of such experience, unlike their previous year's predecessors.

Further, the OSCAR teachers had had the benefit of another year's experience, and the possibie morale problems associated with the OSCAR program (Sewell & Dornseif, 1974, p. 22-23) appeared largely to have been resolved.

The second-year evaluation, then, may be considered more critical than that of the first year. With respect to the eighth grade students, analyses of their achievements and performances would more adequately provide an evaluation of the open program's total and general value; those students had originally been simply randomly assigned to the program. The second year's seventh-graders,

having been randomly assigned to subgroups within the larger groups, would, hopefully, provide a basis for evaluation of each program's special advantages for each of these subgroups. In practice, however, due to attrition and absences at testing times, the number of students in each of these subgroups is typically too small to permit specific subgroup comparisons; hence, most of the following analyses will deal with the seventh grade students as if they had been randomly assigned to each program without reference to subgroup chracteristics. These differences between seventh-graders and eighth-graders, nevertheless, recommend that data for the two grades be separately considered.

All of the analyses and results reported here are based upon year-end data only: measurements secured during May and June of the program's second year. Analyses based upon certain mid-year data of the second year have previously been reported (Dornseif et al., 1974).

Results: Absolute Analyses

As before (Sewell & Dornseif, 1974), two varieties of analyses were conducted: absolute and relative. The absolute analyses are concerned solely with absolute differences between the OSCAR and Control groups. These analyses accept some fundamental, underlying differences between the two programs and seek to establish not the nature of such differences but the ways in which such differences influence student outcomes. Hence the primary focus of these analyses is outcome measurements (the dependent variables) as functions of the two programs (the primary independent variable). Other independent variables of interest are grade level (seventh vs. eighth) and sex (male vs. female). The fundamental design of these analyses, then, is that of a 2 x 2 x 2 MANOVA (or, when appropriate, ANOVA). The results of each analysis are summarized briefly below.

Stanford Achievement Scales. On the basis of previous experience, only

four of the Stanford Achievement scales were administered at year's end: Paragraph Meaning, Arithmetic Concepts, Arithmetic Computation, and Science. The achievement data employed in this analysis were grade equivalents. The multivariate F for Group (OSCAR vs. Control) was not significant ($F = 1.867$, $df = 4/208$, $p > .05$). As would be expected, the multivariate F for Grade was significant ($F = 2.810$, $df = 4/208$, $p < .03$); subsequent univariate analyses showed the two grades to differ significantly only in Arithmetic Computation ($F = 1.222$, $p < .04$) and Science ($F = 8.400$, $p < .005$), and in both areas eighth-graders out-performed seventh-graders.

The multivariate F for Sex was significant ($F = 7.011$, $df = 4/208$, $p < .0001$). The subsequent univariate Fs were significant for Paragraph Meaning ($F = 4.409$, $p < .04$), Arithmetic Concepts ($F = 10.801$, $p < .001$), and Arithmetic Computation ($F = 3.953$, $p < .05$). In each case, the better performance was achieved by female students.

The multivariate Fs for all two-way interactions (Group x Grade, Group x Sex, Grade x Sex) failed to attain statistical significance. The three-way interaction (Group x Grade x Sex), however, was statistically significant (multivariate $F = 2.926$, $df = 4/208$, $p < .03$). The univariate Fs for all of the Stanford scales were significant, and the patterns were as would be expected from the preceding main effects analyses.

It should be noted that, in general, cell means for all of these dependent variables exceeded national norm grade equivalents, and a number of these cell means substantially exceeded national norms.

Learning Environment Inventory. As a measure of student attitudes, the first-year study had employed The School Inventory, a 100-item inventory developed by Bell and yielding a single score. For both practical and theoretical reasons, use of this instrument was not continued during the second-year evaluation. As has been previously reported (Dornseif et al., 1974, pp. 14-18), the

Learning Environment Inventory, developed by Walberg and Anderson (Anderson, 1973) was selected as a measure of students' attitudes: first, because of the face validity of its multi-dimensional approach; second, because it has been employed in a number of studies which appear to support claims of reliability and validity (Anderson, 1973); and third, because a majority of the audience to whom this report is directed can be expected to have some familiarity with the instrument. Preliminary data, secured by mid-year administration of the LEI, appeared to demonstrate its usefulness in this evaluation program, although the authors maintained some concern over the appropriateness of its current application.

Near the end of the year, the LEI was again administered to 103 OSCAR students and 107 Control students. Due to the limited capacity of the scoring computer, only 98 of the 105 items were scored; the seven items of the Diversity scale were eliminated since the reported reliability of these items was lowest.

Scale scores for the remaining 14 scales were analyzed through the study's basic MANOVA design. Of the seven multivariate Fs, four were statistically significant: Group, Sex, Grade, and Group x Sex. The results of subsequent univariate analyses are summarized below.

The multivariate F for Group was 9.195 (df = 14/189, $p < .0001$). The OSCAR mean was greater than the Control mean for three scales: Cliqueness (F = 32.749, $p < .0001$), Disorganization (F = 4.854, $p < .03$), and Democratic (F = 7.896, $p < .006$). The Control mean was greater than the OSCAR mean for four scales: Formality (F = 4.528, $p < .04$), Speed (F = 19.402, $p < .0001$), Favoritism (F = 9.411, $p < .003$), and Apathy (F = 4.150, $p < .05$).

The multivariate F for Sex was 2.060 (df = 14/189, $p < .02$). The mean for females was greater than the mean for males in the case of three scales: Cohesiveness (F = 14.104, $p < .001$), Environment (F = 14.080, $p < .001$), and Goal

Direction ($F = 6.178$, $p < .02$). Scale means were greater for males than for females in three other cases: Favoritism ($F = 5.136$, $p < .03$), Disorganization ($F = 6.403$, $p < .02$), and Apathy ($F = 11.049$, $p < .01$).

The multivariate F for Grade was 1.813 ($df = 14.189$, $p < .04$). Only the Disorganization scale showed univariate significance ($F = 4.337$, $p < .04$); here the eighth grade mean exceeded the seventh grade mean.

The multivariate F for the Group x Sex interaction was 1.959 ($df = 14,189$, $p < .03$). The scales contributing to this significance were Cohesiveness ($F = 10.642$, $p < .002$), Formality ($F = 6.546$, $p < .02$), Environment ($F = 14.828$, $p < .001$), Goal Direction ($F = 5.836$, $p < .02$), Disorganization ($F = 14.884$, $p < .001$), and Apathy ($F = 4.926$, $p < .03$). Directional differences in means are evident from the preceding main effects discussions.

Piers-Harris Children's Self Concept Scale. As noted previously, analysis of first-year data provided by this instrument showed Control students to have significantly higher self-concept scores than OSCAR students; this finding and its implications were sources of considerable concern. Hence the use of this instrument was continued in the second-year evaluation.

This scale yields a single score; the data were, therefore, ANOVA processed. None of the main effects was found significant: Group ($F = 2.68$, $df = 1/209$), Grade ($F = 1.37$), or Sex ($F = .21$). Of the two-way interactions, only Group x Grade was significant ($F = 6.30$, $p < .05$); the F for Group x Sex was 1.99, while the F for Grade x Sex was .31. The three-way interaction, Group x Grade x Sex was not significant ($F = .09$).

Subsequent analyses showed this significant Group x Grade interaction to be solely the result of differences between the seventh grade OSCAR students and the other groupings. Group means are as follows: seventh grade OSCAR, 59.72; seventh grade Control, 53.25; eighth grade OSCAR, 53.36, eighth grade Control, 55.70.

The eighth-graders of the current study had completed the Self Concept Scale as seventh-graders the previous year, so the group means for these students were compared. While the mean score for the Control students had remained virtually unchanged (increasing slightly, from 55.44 to 55.70), the mean for the OSCAR students had shown a greater increase (from 51.79 to 53.36). Analysis, however, showed these changes to be statistically non-significant, even when sex was considered. The general patterns in these year-to-year changes are seen in the following table.

MEAN SELF CONCEPT SCORES

	1974	1973
Seventh grade OSCAR	59.72	51.79
Seventh grade Control	53.25	55.44
Eighth grade OSCAR	53.36	51.56
Eighth grade Control	55.70	57.07

Although analyses of seventh grade subgroups had been intended, the number of students in each subgroup was finally too small, and simultaneously too disproportionate between subgroups, to permit a reasonable statistical analysis. It should be noted, however, that the mean score of each OSCAR subgroup exceeded the mean of the corresponding Control subgroup. Further, within each major group (OSCAR and Control), the highest mean self concept scores were recorded by the "Extraversion" subgroup, and the lowest mean score by the "Introversion" subgroup. Somewhat smaller mean score differences between subgroups were noted in the Control group than among OSCAR students; that is, the Control students appeared generally more homogeneous in self concept than did the OSCAR students.

Results: Relative Analyses

As had been done previously (Sewell & Dornseif, 1974), teachers associated with both programs were asked to provide individual ratings of each

student. In the case of the OSCAR program, each of the four teachers rated all of the program's students; students in the Control group, however, were widely dispersed throughout the Upper Grade Center, and no single teacher was acquainted with all these students; hence ratings of the Control students were made by quite a large number of teachers, each rating a relatively small number of students.

Teachers were asked to rate each student on four dimensions: Attitude, Knowledge, Skills, and Sociability. The following behavioral definitions of these rating categories were provided:

> Attitude: Student displays positive attitudes toward school, teachers, other school personnel, and other students.
>
> Knowledge: Student demonstrates mastery of academic content appropriate to his/her age, grade level, and apparent ability.
>
> Skill: Student demonstrates application of academic content within school and displays ability to apply academic content in non-academic settings.
>
> Sociability: Student demonstrates respect for the rights and feelings of others and demonstrates ability to work effectively and cooperatively with others.

Using a five-point scale, teachers were asked to evaluate each student independently of other students and to use each rating dimension independently of the others. Previous experience with this rating process suggests that the four dimensions probably constitute two rating factors: one described by the Knowledge and Skills categories, and the other by the Attitude and Sociability categories.

The data produced by this rating process, then, consisted of four categorical ratings of each OSCAR and Control student by each of four teachers associated with the respective programs. To eliminate biases introduced by the teacher's idiosyncratic use of the scales, each teacher's ratings of all students was converted into a z distribution, and each individual rating converted

into a corresponding z; these were then summed and averaged across rating categories to provide a mean z rating of each student for each teacher. Finally, the four zs thereby obtained were averaged to provide a mean z rating across categories and across the teachers who had rated the student. The sixteen ratings were compressed in this manner to provide a single, minimally biased evaluation of each student.

The final distribution of mean-z ratings was then divided (separately for each group) at the median. For the OSCAR students, this median was .05, while the median of the Control group was -.06. In each group students achieving ratings greater than the median were arbitrarily classed as "successful" students, while those whose ratings were less than the median were classed as "unsuccessful" students. These dichotomizations provided the bases for all the "success" analyses reported here.

The primary purpose of these "success" analyses is to derive a definition of "success" for each program: such a definition would permit inferences of the program's objectives and would help to identify the distinctive characteristics of each program. As has been previously noted herein, this analytical procedure had prompted the conclusion that during the OSCAR program's first year the objectives of that program had differed little, if any, from the objectives of the traditional program, since profiles of "successful" and "unsucessful" OSCAR students were virtually identical to those of their counterparts in the Control group.

Intercorrelations of Teachers' Ratings. Since all four of the OSCAR teachers had rated the same students, the degree of correlation of each teacher's final z rating with that of each other teacher was calculated. Because of the closeness of the professional relations and the interdisciplinary intent of the OSCAR program, a pattern of relatively high and consistent intercorrelations should be expected. These intercorrelations are given in the following table.

INTERCORRELATIONS OF TEACHERS' RATINGS OF OSCAR STUDENTS

	Language Arts	Social Studies	Math	Science
Language Arts		.632	.577	.608
Social Studies			.671	.697
Math				.684

These correlations are generally higher and far more inter-consistent than those encountered at the end of the program's first year, at which time the range was from .395 to .861 and the mean r was .524 (Sewell & Dornseif, 1974). The consistency of these correlations indicates a substantially greater unanimity of the teachers' perceptions.

Since ratings of the Control students were derived from a large number of teachers, each of whom rated only a few students, it was deemed impractical to attempt intercorrelations of ratings by those teachers.

Grade and Sex Factors in Teachers' Ratings. As has been previously noted, the first-year evaluation indicated that teachers' ratings were apparently strongly influenced by the grade level and sex of the student, such that in the OSCAR program a seventh grade girl was significantly more likely to be rated "successful." Hence a similar analysis was conducted with second-year data. Means of the z ratings are given in the following table.

MEAN z RATINGS OF STUDENTS BY TEACHERS IN OSCAR AND CONTROL GROUPS

	OSCAR		CONTROL	
	Male	Female	Male	Female
7th Grade	.06 (n=31)	.02 (n=35)	.30 (n=27)	.32 (n=32)
8th Grade	-.35 (n=22)	.39 (n=26)	-.28 (n=24)	.02 (n=25)

These data were employed in a three-way analysis of variance, the results of which are summarized in the following table.

ANALYSIS OF VARIANCE: MEAN z RATINGS OF STUDENTS BY TEACHERS

Source	SS	df	F
Grade (G)	2.226	1	4.730*
Sex (S)	2.881	1	6.014*
Program (P)	.248	1	.517
G X S	3.779	1	7.889**
G X P	2.801	1	5.847*
S X P	.258	1	.538
G X S X P	.826	1	1.724
Error	102.568	214	
Total	115.627	221	

*p < .05
**p < .01

The two programs do not differ significantly in this analysis, of course, by virtue of the separate z transformations of the original ratings. Examination of the preceding table of cell means shows the directionality of the significant differences. Seventh-graders receive higher ratings than eighth-graders; girls receive higher ratings than boys; eighth grade boys receive the lowest ratings; and differences in ratings are most striking among Control group boys.

Grade and Sex Factors in "Success" and "Failure." The preceding analysis was not concerned with the roles of grade and sex in the student's assignment to the "successful" and "unsuceesful" comparison groups previously described. The significance of these factors was tested through a series of chi-square analyses, separately performed for each sex, each grade level, and each group. Similar analyses of first-year data, it will be recalled, had shown seventh grade OSCAR girls to be disproportionately "successful," and this finding, in association with other findings, had suggested a higher emphasis upon quiet and orderliness in the OSCAR program than appeared in the traditional program. Relevant second-year data are shown in the following table.

NUMBERS OF STUDENTS RECEIVING MEAN z RATINGS
AS "SUCCESSFUL" OR "UNSUCCESSFUL"

		OSCAR		CONTROL	
		Male	Female	Male	Female
7th Grade	Successful	16	14	11	21
	Unsuccessful	15	21	16	11
8th Grade	Successful	7	20	8	14
	Unsuccessful	15	6	16	11

In OSCAR, sex alone is not a significant contributor to "success" (χ^2 = 1.72, df = 1, $p > .05$); nor is grade level alone (χ^2 = 1.28, df = 1, $p > .05$). Considered simultaneously, however, sex and grade level are significant predictors of "success" (χ^2 = 10.32, df = 3, $p < .02$); despite this significance, the strength of this association is not particularly impressive: the Goodman-Kruskal index of predictive association (Hays, 1963, pp. 606-610) is .122. As the data of the preceding table show, OSCAR seventh grade girls are disproportionately "unsuccessful" and eighth grade girls are disproportionately "successful; somewhat opposite trends hold for the seventh and eighth grade boys, but to a lesser extent.

In Control, however, the simple relation of sex to "success" is significant (χ^2 = 6.26, df = 1, $p < .02$), while grade level is not (χ^2 = .92, df = 1, $p > .05$). In this group, boys of both grade levels are disproportionately "unsuccessful," while girls are disproportionately "successful."

Stanford Achievement Scales. Three different analyses of Stanford scores (Paragraph Meaning, Arithmetic Concepts, Arithmetic Computation, and Science scales) in relation to "success" groupings were conducted: OSCAR "successful" vs. OSCAR "unsuccessful"; Control "successful" vs. Control "unsuccessful"; and OSCAR "successful" vs. Control "successful." All took the form of multivariate analyses of variance.

In the comparison of OSCAR "successful" vs. "unsuccessful" students,

the multivariate F was significant (19.767, df = 4/99, $p < .0001$); subsequent univariate analyses showed significant differences for all four of the scales, the higher mean scores in each case having been achieved by the "successful" students.

In the comparison of Control "successful" vs. "unsuccessful" students, the multivariate F was also significant (10.969, df = 4/98, $p < .0001$), and the univariate Fs for each of the scales was significant. Again the higher scores were attained by the "successful" students.

In comparing "successful" OSCAR students to "successful" Control students, the multivariate F was not statistically significant (1.709, df = 4/103, $p > .05$). Scale means for the OSCAR students were in all cases greater than those of the Control students.

Learning Environment Inventory. As before, only 14 of the 15 scales of this inventory were employed. The same multivariate analyses were conducted as in the case of the Stanford Achievement Scales.

In the comparison of OSCAR "successful" to "unsuccessful" students, the multivariate F was significant (2.617, df = 14/85, $p < .003$). Univariate analyses showed "unsuccessful" means to exceed "successful" means on three scales: Speed, Favoritism, and Satisfaction.

The comparison of Control "successful" and "unsuccessful" students produced a significant multivariate F (2.437, df = 14/83, $p < .006$). "Successful" students achieved higher means on the following scales: Cohesiveness, Environment, Goal Direction, Apathy, and Disorganization. "Unsuccessful" students scored higher on Friction, Democratic, and Satisfaction.

The comparison of "successful" OSCAR students to "successful" Control students yielded a significant multivariate F (5.570, df = 14/88, $p < .0001$). OSCAR students achieved higher mean scores in Difficulty and Democratic, while Control students' means were higher on the Formality, Speed, Favoritism, and

Cliqueness scales.

Piers-Harris Children's Self Concept Scale. Mean scores on this instrument for each cell of the analytical design are shown in the following table.

MEAN SELF CONCEPT SCORES

	OSCAR		CONTROL	
	7th Grade	8th Grade	7th Grade	8th Grade
Successful	63.10 (n=28)	57.37 (n=27)	56.63 (n=30)	59.85 (n=21)
Unsuccessful	56.74 (n=35)	47.33 (n=18)	49.20 (n=24)	52.33 (n=24)

Because of the previously noted disparities in distribution among cells of the larger design, sex was not considered in the analysis of these data.

The results of a three-way analysis of variance of self concept scores are provided in the following table. The "Rating" variable simply refers to the "successful" vs. "unsuccessful" categorization.

ANALYSIS OF VARIANCE: SELF CONCEPT SCORES

Source	SS	df	F
Program (P)	324.56	1	2.361
Rating (R)	2492.50	1	18.135***
Grade (G)	222.69	1	1.620
P X R	13.26	1	.096
P X G	946.30	1	6.885**
R X G	120.11	1	.873
P X R X G	718.69	1	5.229*
Error	27351.11	199	
Total	32189.22	206	

*p < .025
**p < .01
***p < .0005

In order to explore contributions to the obviously strong relation between teachers' ratings and students' self concept scores, a series of Pearson rs were computed. Differences in these coefficients will, of course, aid in understanding the interaction patterns reported above. These correlations are reported in the following table.

CORRELATIONS BETWEEN MEAN z RATINGS AND SELF CONCEPT SCORES

	OSCAR	CONTROL
Total Group	.356 (n=108)	.317 (n=99)
7th Grade	.397 (n=63)	.267 (n=54)
8th Grade	.325 (n=45)	-.456 (n=45)
Males	-.347 (n=49)	-.192 (n=47)
Females	.350 (n=59)	.385 (n=52)

DISCUSSION

As noted previously, the two varieties of analyses reported here -- absolute and relative -- had two different but complementary purposes. The absolute analyses, of course, were intended simply to discover any absolute differences in outcomes of the two programs. The relative analyses were intended to explore and define differences in the natures of the two programs. Two educational programs may yield quite different outcomes despite any essential differences in the programs themselves; and such different outcomes might then be attributed to differences in the students, in the teachers, in peculiar interactions of students and teachers, in differing physical environments, etc., etc. Outcome differences stemming from inherent differences in students participating in each program have specifically been eliminated from this evaluation design by random assignment of students to the two programs; nevertheless, no similar controls were possible or practical with respect to the programs' teachers or the physical environment in which each program is conducted. Hence a design limited to absolute analyses of student outcomes necessarily begs the question of whether the programs are truly different.

Many innovative programs fail to achieve their stated objectives, while others achieve certain objectives only at the expense of other objectives. An early and very serious concern of those associated with the OSCAR program was whether traditional educational objectives might not be relegated to a subordinate status in the pursuit of wider, non-traditional objectives. Hence academic achievement of the two groups has been measured at every stage of the evaluation. And it gratifying to find that whatever else the OSCAR program has achieved, the high educational standards of the District have not been sacrificed: the academic achievements of OSCAR students do not differ from those of students in the traditional program; indeed, if anything, there appears to be a slightly higher (statistically non-significant) level of academic achievement in the OSCAR program.

Overall, the findings of the Learning Environment Inventory are equally gratifying. OSCAR students perceive their environment as more clique-ish, more disorganized, and more democratic; in view of program and physical environmental differences in the two programs, such differences are quite appropriate. The perceptions of disorganization can easily be attributed to the sheer number of students housed in a single, large classroom. Higher Cliqueness scores probably represent both a normal clustering of students within the physical environment and the teachers' efforts to establish small teaching and learning groups. That clique-ishness is not necessarily adversely evaluated is substantiated by the students' perceptions of the environment as more democratic.

Complementing these interpretations, Control students perceive their environment as more formal, more speed-oriented, more tainted by favoritism and more characterized by apathy. The clique-ishness of the OSCAR program is paralleled by the favoritism found in the Control program. That is, the existence of cliques within the OSCAR program is apparently not a product of teacher favoritism.

The first-year finding of higher self-concept scores for Control students was a source of particular concern. This trend was clearly reversed during the OSCAR program's second year. It well may be that the OSCAR teachers' increasing experience and confidence were significant factors in this reversal, since it is most notable amongst the seventh-graders of that program and constitutes the major difference in self-concept scores during the program's two years of operation.

The major conclusion of the first-year evaluation of the OSCAR program was that the absence of student outcome differences was most likely due to the absence of differences in the two educational programs; that is, the OSCAR program evidently had not at that time established a separate identity. This conclusion was based upon an inability to find consistent and significant differences through the various relative analyses reported (Sewell & Dornseif, 1974). The data of the present report, however, indicate rather clearly that such a distinctive identification has been established. A much greater uniformity of perceptions of their students by the OSCAR teachers suggests that a consistent basis for such perceptions has been achieved. Similarly, intercorrelations of Stanford Achievement scores within the OSCAR program are consistently higher than similar intercorrelations within the traditional program -- and, indeed, consistently higher than similar intercorrelations derived from the national standardization sample for these tests. Taken together, these two factors indicate a substantial degree of interdisciplinary agreement and coordination, evidently a necessary ingredient of an OSCAR-type program.

That the two programs differ is also attested by the differing relations of teachers' ratings to student characteristics in the two programs. In the traditional program, this relation is fairly simple: girls are significantly more likely to receive higher ratings than boys. In OSCAR, however, ratings of girls appear to increase substantially between the seventh and eighth grades,

while ratings of boys show complementary decreases. (Incidentally, the consistently lower ratings of eighth grade boys has been repeatedly observed and has been verbally confirmed by other educators; this evaluative phenomonon appears to be widespread and may be related to the conset of male adolescent behavior patterns.)

Hence in terms of these ratings alone, the two programs appear to have distinctive identities. The remaining analyses were intended to establish the nature of these differences. It is quite clear from the data that academic achievement is equally important to "success" in both programs.

In terms of the scales of the Learning Environment Inventory, "successful" OSCAR students perceive their environment as characterized by difficulty and democracy, while their counterparts in the Control group find their environment characterized by formality, speed, favoritism, and clique-ishness. Within the OSCAR program "success" seems to be related to an environment characterized by less speed, less favoritism, and less satisfaction. Within the traditional program "success" is related to an environment characterized by more cohesiveness, a more enjoyable physical environment, greater goal direction, more apathy, more disorganization, less friction, less democracy, and less satisfaction. The relation of "success" to less satisfaction in both programs poses some challenges to further research.

A conceptual dissatisfaction with the nominal scales of the LEI (and their sheer number) led to a still-continuing factor analytic evaluation of this instrument. Using item scores as the raw data of this analysis of 98 items, combined-group data (that is, derived from both OSCAR and Control students) failed to yield any easily understandable factors, despite various types of rotations. When, however, data from each group were separately analyzed, two or three (not 14) reasonably coherent factors did emerge from each analysis, and these factors were conceptually quite different -- nearly complementary -- for

each group. The existence of these disparate factor structures within each group can be taken as further evidence of the distinctiveness of each program. Details of these factor analyses will be reported at a later time, but in the meantime some interpretive caution with respect to the LEI scales seems appropriate.

Analyses of the self-concept scale scores show a high self-concept to be closely related to "success" in both programs. The three-way interaction (program x rating x grade) noted in these analyses reflects both the generally high self-concept scores of OSCAR seventh-graders, and the anomalous negative correlations between self-concept and teachers' ratings amongst males in general and amongst eighth grade Control students in particular.

Unfortunately, as noted previously, the small numbers of students in each program's seventh grade subgroupings precluded detailed statistical analyses. Inspection of subgroup means, however, indicates data trends generally consonant with whole-group results. To some extent the data suggest that the five nominal subgroups (Random Assignment, Underachiever, Introvert, Extravert, and Teacher Nominated) may be functionally described as really only two subgroups: Introvert (including, generally, the Underachiever and Introvert subgroups) and Extravert (including, generally, the Random Assignment, Extravert, and Teacher Nominated subgroups). Typically, the Introvert group has, of course, lower mean self-concept scores and lower teachers' ratings, while opposing tendencies are evident in the data of the Extravert group.

Overall, the results of this second-year evaluation demonstrate that the OSCAR program has established a separate and distinctive identity, that that identity broadly conforms to commonly expressed characterizations of "open education," and that student outcomes of this program are in the desirable directions with respect to both academic and non-academic achievements. The somewhat negative tendencies of the program noted as a result of the first-year

evaluation have been either eliminated or substantially ameliorated. The OSCAR program appears to have established its educational feasibility and value, and its continuation (and continued evaluation) has been recommended.

Finally, the differences between the results of the first-year evaluation and the present evaluation strongly support the need for (1) early planning of evaluation, (2) development and continued use of an appropriately critical evaluation model, and (3) continuous and continued evaluation, rather than premature acceptance or rejection of a program on the basis of premature or hasty evaluation.

References

Anderson, G. J. The assessment of learning environments: a manual for the Learning Environment Inventory & the My Class Inventory. Halifax, N. S., Canada: Author, 1973.

Dornseif, A. W., Gross, S., & Sewell, A. F. Multivariate evaluation of student selection strategies in open and traditional education. ERIC Document 093 930. (1974)

Hays, W. L. Statistics. New York: Holt, Rinehart & Winston, 1963.

Sewell, A. F., & Dornseif, A. W. Controlled multivariate evaluation of open and traditional education at the junior high school level. ERIC Document 074 144. (1973)

Sewell, A. F., & Dornseif, A. W. Controlled multivariate evaluation of open and traditional education at the junior high school level: the first-year report. ERIC Document 090 299. (1974)